日本の「AI事業者ガイドライン（第1.2版）」：実務対応の最前線

第1.2版の核心：定義の拡張とリスクの再定義

DEFINITION: AIエージェント

自律的に行動するAIシステム

特定の目標達虚のために環境を感知し、自律的に権限を行鑒する（自勤特法、迷金等）システムを定義し、責任分配の明確化を求めています。

DEFINITION: フィジカルAI

物理世界に直接作用するAI

アクチュエータを介して現実世界に働きかけるAIを揺し、サイバー登撻だけでなく人身・説覧への危害防止を最優先事項としています。

KEY_FINDING:
「学習」「推論」の定義拡張

RAGや外部知識連携も管理対象へ

リスク速をモデル単体ではなく、推課的に参照する外部データやAPI、プロンプト法入のリスクまで含めて共通理解としています。

外部API
プロンプト
外部データ
自律的AIシステム
プロンプト
アクチュエータ

業界別：対応コストと追加負担の目安

STATISTIC: 製造業（ロボット・検査）

追加負担：約130万〜820万円

安全性検弱や停止基準の實定のため、年間3〜12人月の追加工数が想定されます。

STATISTIC: 金融業（与信・不正検知）

追加負担：約265万〜1,635万円

公平性の硬保や人間判断の介在、高度な記提保時のため、6〜34人月の工数が必要になる可能性があります。

STATISTIC: 医療（診断支援・事務）

追加負担：約180万〜1,226万円

説明責任、プライバシー、誤作動の影響評価に4〜18人月の工数が想定されます。

実務担当者が今すぐ取るべき3つのアクション

1. AIの棚卸しと役割マッピング

自社の立ち位置と責任範囲の確定

開発者・提供者・利用者のどの立場かを整理し、SaaS利用や業務委託を含めて責任の「相手」を確定させます。

2. トレーサビリティの実装

「監査可能」なログと履歴の確保

データ出所、意思決定プロセス、外部API呼出、視聞係値の履歴を保存し、事動物の原因究所を可能にします。

3. AI特有のインシデント対応

新たな脅威へのガードレール設置

プロンプトインジェクションやDoS攻撃、誤動作を確定したセキュリティ運用を、設計段階から組み込みます。

説明責任（アカウンタビリティ）のフロー

ステークホルダー説明の設計

「誰に、何を、どこまで」説明するか

単なる技術解説ではなく、金融・医療などの処理に応じた説明レベルのテンプレート化が推奨されます。

サプライチェーン全体の統制

ベンダ管理と責任分界の明確化

調達要件にトレーサビリティの提供を明記し、委託先との契約で責任の所在を事前に定義する必要があります。

国際比較：日本のガイドラインの立ち位置

項目	日本 (1.2版)	EU (AI Act)	米国 (NIST RMF)
法的性格	ガイドライン（自主的取組）	規制（法的義務・罰則あり）	フレームワーク（任意）
リスク分類	重要者によるリスクベース	高リスク等の法的枠組み	機能・組織能力による管理
生成物の識別	推奨（透かし等）	透明性義務あり	フレームワーク内での推奨